

Review of Sentiment Analysis: An Hybrid Approach

Asoshi Paul Anule^{1*}, Moshood Abiola Hambali² & Chukwudi Jennifer Ifeoma³

^{1,2}Department of Computer Science, Federal University Wukari, Taraba State, Nigeria.

³Department of ICT, Federal University Wukari, Taraba State, Nigeria.

Corresponding Author (Asoshi Paul Anule) - asoshipaulanule@gmail.com*



DOI: <https://doi.org/10.46431/MEJAST.2022.5405>

Copyright © 2022 Asoshi Paul Anule et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article Received: 19 November 2022

Article Accepted: 17 December 2022

Article Published: 23 December 2022

ABSTRACT

Sentiment analysis is acknowledged as detecting thoughts used from field content features additionally it's recognized while one linked to the main parts of standpoint extraction. Through this type of process, we will be able to discover if a movie script is positive, negative, or natural. Using this research, a feeling examination is executed along with calvados data. The text message sensation analyzer combines organic and natural language processing (NLP) and even machine studying techniques to provide measured assessment rankings to be able to entities, subjects, themes, and groups in a term or key phrase. Inside expressing feelings, the particular polarity of calvados written content reviews can always be graded for the damaging to good range utilizing the education algorithm. The certain current decade presents seen substantial improvements in artificial brains; along with the device mastering revolution offers converted the complete AI sector. In the end, unit learning techniques include grown to always be an important aspect of any design and style in today's absorbing world. However, this ensemble of researching techniques promises for anyone who is part of motorization using the removal of common regulations for textual written content message and sentiment category activities. This kind of particular thesis has to style and carry out a good superior functionality matrix employing ensemble studying intended for sentiment category while well as software. With this paper, we possess analyzed the well-known techniques adopted within the classical Emotion Analysis problem associated with analyzing Elections evaluations like; Support Vector Machine (SVM) and Linear Regression (LR) for the effective detection of sentiments from the dataset obtained from the Kaggle machine learning repository.

Keywords: Sentiment analysis; Text message sensation analyzer; AI sector; Support vector machine (SVM); Linear regression (LR).

1. INTRODUCTION

1.1. Background of the Study

Global evolution in technology and sociocultural behavior has dynamically revolutionized the world. As a result, people share their ideas, thoughts, beliefs, opinions, and decisions in real-time on different platforms such as Twitter, Facebook, and LinkedIn, amongst others (Saif, 2017). However, gaining insights from these textual data may be a herculean task, consequentially due to the massive volume of information from a large population and the complexity associated with quantifying the text data for making the appropriate and beneficial decision during its modeling (Saif, 2017).

Sentiment analysis, also referred to as opinion mining, is the field of study that qualitatively analyzes opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such; as products, services, organizations, individuals, issues, events, topics, and their attributes and thus classifies the review under a predefined polarity (Dey *et al.*, 2019). As a field of research, it is closely related to (or can be considered a part of) computational linguistics, natural language processing, and text mining. Proceeding from the study of affective state (psychology) and judgment (appraisal theory), this field seeks to answer questions long studied in other areas of discourse using data mining and computational linguistics (Birjali *et al.*, 2021). Nowadays, in the conductance of sentiment analysis, clear and straightforward instructions are crucial for obtaining high-quality annotations. Nevertheless, sentiment reviews are unstructured, with diverse meanings applicable to the lexicon (Agaian and Kolm, 2017). This ambiguity attached to the corpus makes the classification of sentiment polarity difficult. Most often, sentiment polarity for a particular text feature is dichotomized into positive and negative (Phan *et al.*, 2020).

A document containing several opinionated statements typically contains mixed polarity, thus increasing the complexity of sentiment analysis (Phan *et al.*, 2020). It is noteworthy that sentiment, whether positive or negative, does not happen in a vacuum; rather than obsess over a one-off compliment or complaint, sentiment considers an individual's emotional voicing (Cachola *et al.*, 2018). Therefore, the toxic nature of sentiment can detrimentally affect one's psychological living with an inferiority attachment to its victim (Cachola *et al.*, 2018). Consequentially, it causes people to stop expressing themselves and hence give up on seeking different opinions to make the best decision. For a business, firms, governments, organizations, Etc., it might result in a loss in their financial expectation.

Nevertheless, the importance of sentiment analysis cannot be over-emphasized as its application and impact span diverse fields and domains (Hasan *et al.*, 2018). Although sentiment analysis has diverse demerits, organizations, companies, agencies, and governments still leverage sentiment analysis to gain insight that can enhance efficient and effective decision-making.

Due to several sentiment analyzers, there have been many attempts by researchers to come up with sentiment analysis methods capable of efficiently and effectively detecting sentiment from a stream of textual opinion. Most of these sentiment analysis methods presented by the authors have focused on the sentiment analysis approach using various deep learning and machine learning methods (Al-Smadi *et al.*, 2017; Abdi *et al.*, 2019; Yadav and Vishwakarma, 2020). More specifically, deep learning approaches such as Convolutional Neural Network (CNN) were applied by (Deng *et al.*, 2022) to achieve high-precision text sentiment analysis, a Bidirectional Encoder Representations Transformers (BERT) was applied by (Ray *et al.*, 2020). Considering the Machine learning model, (Fayyouni & Idwan, 2021), amongst others, have applied Naïve Bayes (NB), J48, and Logistic Regression (LR) classifiers to predict the polarity of the collected tweets from an Arabian tweet. All of these signify the viabilities of machine learning and deep learning model in sentiment analysis.

However, despite many efforts conducted in the sentiment analysis, the existing approaches still suffer from a high false-positive rate in detecting sentiment as positive or negative polarity. Moreover, the research on using Machine and Deep Learning methods for sentiment analysis is currently in its problematic stage with enormous demand for feasible solutions. Therefore, this paper proposes the use of two distinct machine learning models, namely, the Support Vector Machine (SVM) and Linear Regression (LR), for the effective detection of sentiments from the dataset obtained from the Kaggle machine learning repository.

2. LITERATURE REVIEW

2.1. Theoretical Framework/Review

Sentiment analysis is, at the moment, regarded as among the list of exciting research topics in natural language processing (NLP). Sentiment analysis mainly aims to spot user opinions and emotions through written content.

The concept of sentiment analysis lies in managing emotions, opinions, and subjective texts (Yeole *et al.*, 2015). Sentiment analysis provides information about public opinion when analyzing various tweets and reviews. This validated tool predicts many important scenarios, such as movie control office performance as well as general elections (Heredia *et al.*, 2016). Public evaluation evaluates a specific entity, such considering that a person, items,

or location. It can be on various websites, such as Amazon and Yelp. Opinions can be negative, positive, or neutral. The objective with regards to sentiment analysis, it sometimes automatically discovers typically the meaning full life long user reviews (Luo & Cao 2016). The particular advantages of emotion exploration are driven by typically the growing need to be able to review and construct undetectable information coming from interpersonal media inside unstructured data (Haenlein & Kaplan, 2010).

Sentiment analysis systems collect research from unorganized, unstructured text that organizations collect coming from online sources for example email, blog posts, support tickets, web chats, social media channels, forums, and comments. Beneficial for Algorithms to replace manual data processing with rule-based, automatic, or hybrid process implementations. Rule-based systems perform sentiment analysis based on pre-defined lexicon-based rules, while automated systems use device learning techniques in order to study from info. Hybrid sentiment analysis can be a combo of equally approaches.

Sentiment analysis is becoming increasingly crucial for exploring ever-increasing opinions on social media and other websites at an unprecedented rate. The enormous information explosions in the telecommunications, aviation and alternative markets of recent years have taken control of all this enormous amount of information and analysis has been done in traditional ways. Hence, scientists and researchers are very enthusiastic. We have developed an efficient technology. These require sentiment analysis as a way to process the data and determine their own polarity to make the right selection.

2.2. Sentiment Analysis Process Steps

Sentiment analysis involves five data processing steps. The data acquisition, text editing, emotion detection, emotion classification, and output presentation (Alessia, 2015) are shown below.

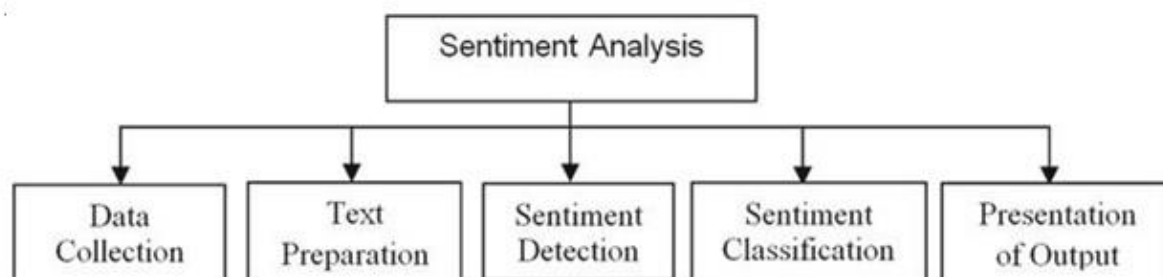


Figure 2.1. Sentiment analysis process steps (Alessia *et al.*, 2015)

2.2.1. Data Collection

The particular critical first stage of sentiment study is Data collection. Data are collected originating from user groups, tweeter, Facebook, blogs, along with commercial websites such as amazon.com and alibaba.com. This specific data cannot end up being analyzed by standard scanning, text research, and language running methods.

2.2.2. Text Prepare

Text preparation examines the data before analyzing it. For example, some reviews and conversations on communication sites contain offensive and inappropriate language and are being investigated and prepared for more reliable analysis results. This process sorts out and removes content that is irrelevant to your examination.

The purpose linked with this procedure would likely be to remove spam and inappropriate reviews before they are automatically analyzed. In this case, you can use the part-of-speech (POS) technique to prepare the Text before analysis (Semih, 2014; El-Din, 2015)

2.2.3. Sentiment Detection

Sentiment detection can be seen as a way about actually finding the particular sentiment newline portrayed in an evaluation with the help of system learning techniques or NLP techniques; these are also known as opinion mining (O.M.) new line and sentiment analysis. Sentiment detection consists of examining phrases and sentences extracted from reviews and ideas. All self-expression phrases such as considering that opinions and mistreatment are maintained. Several research studies in this area include various detection methods, such as those (Lakshmish, 2013). Retro evaluation on experiencing analysis aim to polarize a given text by classifying this as positive, negative, or neutral. This categorization need is considered one of the critical limitations of traditional sentiment analysis. Many scholars think belief analysis in its traditional form cannot deal with the down attributes of modern-day expression, as it comes flat for recording objectivity and subjectivity. For example, traditional sentiment examinations are usually insufficient if the range is tasked to sort between false mass media or someone's opinions and facts. Many superior methods test to admit multiple differentiated efficient indications in text messages, which suggest viewpoints through research from the terminology used for self-expression. Such form of methods often targets simultaneously and have matter models. For this reason, serious learning strategies such as convolutional nerve bodily organs networks (CNNs) are always used. CNN's double in experiencing quest for short-form text messaging, as explained in multiple research papers e.g., (Dos *et al.*, 2014), (kale *et al.* 2018), and (Tang *et al.* 2015).

2.2.4. Classification Approach of Sentiment Analysis

Sentiment classification is a text extraction and classification task that aims to classify Text according to the polarity of the opinions contained therein (Pang, 2002).

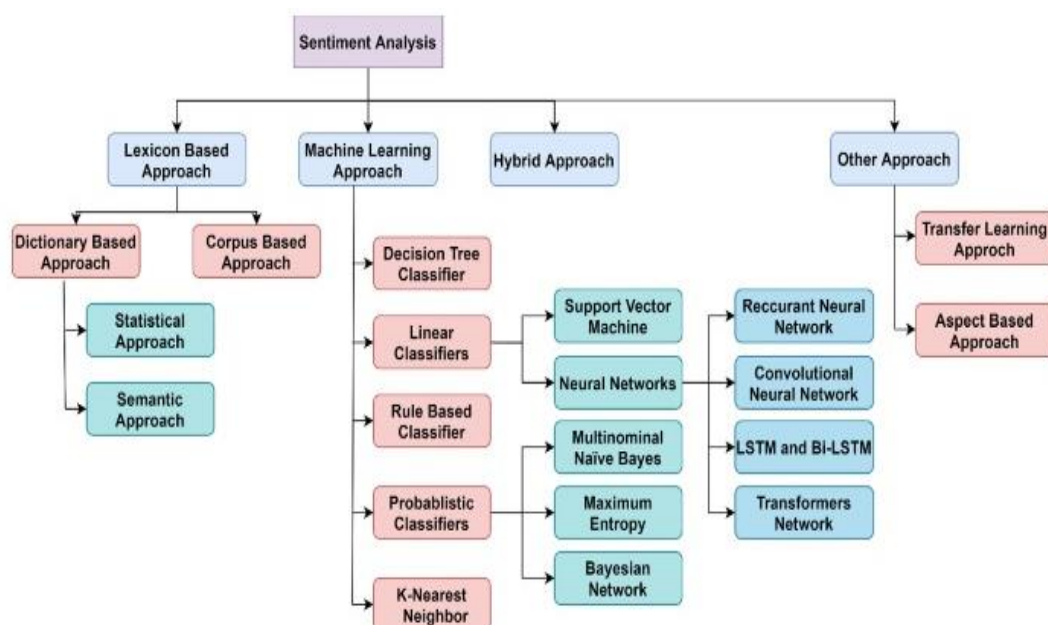


Figure 2.2. Sentiment classification techniques (Alessia *et al.*, 2015)

Positive or negative, good or bad, like or dislike. Emotion classification includes several methods and is divided into three main methods: machine learning, hybrid methods, and dictionary-based approaches (Vaghela, 2016); (Biltawil, 2016). Currently, naive Bayes technology and support vector machines (SVMs) are more common and are used for emotion classification. These techniques can improve the accuracy of tweet classification, such as Naive Bayes. Use Naive Bayes for sentiment analysis tweets in Fast (Ankur, 2016). As a result, sentiment analysis has undergone much research in this area, and in recent years there have been many applications and improvements in sentiment analysis. Sentiment classification techniques are described with a focus on most details and related points and route references. The following figure shows all the techniques used so far for sentiment classification.

2.2.4.1. Lexicon-Based Approach

The strategy uses multiple words to classify a mood; positive words are used for what is needed, and negative words for what is not needed. Therefore, the particular lexicon-based approach depends upon primarily on the search for opinion lexicons used for text analysis. According to the dictionary-based approach, there are two ways. One is a corpus-based approach, and the second is a dictionary-based approach (Aqlan *et al.*, 2019).

Typically most of the lexicon-based method is very functional at the word and has a level of understanding analysis. However, expenses demand any training info. Thus, it could be considered an unsupervised method. However, the critical problem with this procedure is website dependency because words will indeed have multiple connotations and senses; thus, a good word in a sure website name might not take an additional. For example, offered the expression "small" and two phrases "The TV SET display is, in fact, small". Furthermore, "This camera is tiny". the phrase "small" in the first word is negative because people generally favor large displays. However, the second phrase is positive; like the camera is small, it will be easy to bring. This specific issue can be overlooked by introducing any domain-specific sentiment lexicon or using a lexicon variation method. (Sangar *et al.*, 2020) suggested a new genre-level emotion lexicon variation method. Opposite to other variation techniques that use branded data, this new strategy employs unlabeled data to learn the source and the focused website sentiment lexicons. The particular transfer learning techniques can be used to learn new domain-specific lexicons, as in the job of (Sanagar *et al.*, 2020). The particular creators suggested an unsupervised emotion lexicon learning strategy you can use for brand new websites of identical type. Right after learning the polarity seeds, and words from corpora of multiple source websites, the genre-level knowledge uncovered can then be sent to the aimed domains. An additional problem of the lexicon-based approach is the dropped performance compared to the machine learning approach if a huge training dataset is provided. Here are three primary techniques for creating and annotating belief lexicons (Asghar *et al.*, 2019).

2.2.4.1.1. Corpus-Based Approach

The corpus-based approach starts with a list of opinion words and finds other ideas from words in a large corpus to get opinions from a particular direction. In another sense, most methods rely on grammatical patterns in the first list of words of opinion to find other words from the large corpus (Hatzivasiloglu, 2004). Therefore, the first step was to create a seed list and use it with various language constraints so that it could identify other words that contained directions. Two approaches are used to implement the corpus-based approach: the statistical approach and the semantic approach.

i. Statistical Approach: This method gives in many programs related to emotion analysis.. The most famous of these is the one that can detect scoring operations by running a statistical randomization test called the gait test (Kim, 2004).

This sort of method acquires the sentiment alignment of any word in line with the statistics principle. The basic principle of this method is that similar idea words will often have the same belief when they turn out collectively frequently in the same circumstance. As a result, the unidentified polarity of the term is obtained in line with the regularity from the co-occurrence of words that are shown upwards collectively in the same situation. The regularity of co-occurrence is calculated using Turney's method for mutual computer information (Turney *et al*, 2003). A new amount of appropriately used approaches to build sentiment lexicons and perform experience analysis. (Han, *et al*, 2018) advised a brand new domain-specific lexicon for reviewing sentiment evaluation. They will use contributed information to provide conditions with their Atrás tags in the lexicon. The particular authors got a good outcome using the proposed method.

ii. Semantic approach: It provides value to sentiments while relying on more than principles to calculate the affinity and similarity of different words. The basis of this principle is to support the value of words and their sentiments (Mohammad, 2009).

The previous method (also known as the ontology-based method) uses diverse regulations to gauge the particular likeness between phrases and designates the particular same sentiment worth immediately to the actual semantically close phrases (Araque *et al*, 2019). Usually, this method appears way up emotion alternatives, antonyms, and phrases with a similar principle to lengthen a lexicon and perform feeling analysis, much like (Zhang *et al*, 2012). The specialists blended statistical and semantic approaches to suggest Weakness Person, a professional system that finds product weaknesses through Oriental reviews. These folks used the Chinese Hownet (Dong *et al*, 2006) lexicon determine the similarity from Text. Typically suggested professional system exhibited fantastic performance around trial and error results.

2.2.4.1.2. Dictionary-Based Approach

The dictionary-based way offered complete method for the dictionary-based way. In this well-known approach, a small organization of phrases is hand-picked with regarded trends (Miller, 1993; Hatzivassiloglon, 1998). Then plant this organization of phrases by looking inside their graded method corpora word list (Medhat, 2014). The new phrases discovered are delivered to the seed list, and the following repetition begins. The repetitive process keeps forestalling and only stops when there are no new words.

Almost all dictionary-based approaches involve pre-defined set view words collected (Chetviorkin *et al*, 2012); (Kaity *et al*, 2020). Almost all guess powering this way is that word, and expression replacements include the same polarity as the bottom phrase, while antonyms include opposite polarity. Significant corpora like series of word and phrase replacements or word net will be thought about for antonyms and word and expression replacements, following which it is appended to a class or seedling document prepared previously. In the first stage, a principal pair of keywords is accrued physically with the orientation. After suggestions are widened, seek the antonyms and appearance and phrase alternatives in the offered lexical resources (Singh *et al*, 2017); (Ho *et al*, 2014). Then

typically, the words are iterative as part of the file and the file is expanded. Guideline examination or a static correction may be performed during the past levels to ensure this quality. Stefano and Hazel created SentiWordNet three-way in (Baccianella *et al.*, 2010) using the assistance regarding computerized annotations regarding WordNet 3's synsets. Another famous valuable resource, the thesaurus, was performed based on online dictionaries. Generally, the operation of (Park and Kim 2016) suggested a rule-based method for labeling experience paragraphs and keywords in in-text promotion by employing a dictionary-based approach. This strategy is feasible simply for small syndication sizes. Another downside of all lexicon-based approaches (Hajek *et al.*, 2020), such as the dictionary-based approach, is finding view words and keywords specific for every website, as typically, the polarity can convert.

2.2.4.1.3. Manual Method

Guidebook way requires male intervention to annotate the lexicon. A lot of the creation of suffering from lexicons contains various phases, precisely, the sentiment-bearing words and phrases and phrases document and typically the assignment of suffering from labels to these kinds of varieties of words and phrases. This procedure is typically quite time-consuming, high-priced, and even moment consuming, but it supplies a regular and, in many cases, reliable lexicon. A computerized strategy may be suggested while a factor in improving this approach. If this occurs, some guide approach is applied while some sort of benchmarking method alongside lower typically the errors. A fantastic deal of lexicons has been recently created bodily. (Wilson *et al.*, 2005) approach, (Taboda *et al.*, 2011) Made almost the entire Semantic Positioning Online auto loan calculator (SO-CAL) that could be structured in handbook databases.

Professionals, in addition, can use crowd sourcing and even ramification. Crowd sourcing will always be the practice to utilize a lot intended for some type of famous goal online internet sites. For example, (Turney *et al.*, 2013) used Amazon Genuine Turk to generate an expression feeling. (Hong *et al.*, 2013) developed some sort of game called Composition involving Babel to have interaction players to be able to designate an experiencing polarity to words and phrases suitable for building some sort of feeling lexicon.

2.2.4.2. Machine Learning Approach

Tool learning way fix problems related to being able to manage to text class, including syntactic or linguistic features. For example, a dictionary-based approach extracts emotions from Text but relies on an emotion dictionary. A collection of available precompiled emotional terms for machine learning algorithms. It can be divided into reinforcement, unsupervised, and supervised learning (Medhat, 2014).

Resources knowing way are widely-used to form belief polarity (e.g., negative, positive, and neutral) methodized on mentor and test datasets. These techniques can be broken down into supervised learning (Oneto *et al.*, 2016), unsupervised learning (Li *et al.*, 2017), semi-supervised learning (Hussain *et al.*, 2018), and support learning (Rong *et al.*, 2014). Viewed method will be applied when this classification activity has a unique partner of classes; then, when it is difficult to determine it, thanks to a deficiency of branded data, the unsupervised technique can be your case. Throughout area, the semi-supervised strategy can supply unlabeled datasets, which include some proclaimed examples. Most of the strategies of support learning use trial and error components to ensure that the realtor affix to the surrounding atmosphere to get maximum rewards.

Equipment learning strategies can teach domain-specific habits from the written Text, which reasons better category results. However, the condition with these strategies is that they often require extensive training datasets to perform a good performance. Even so, a configured répertoirer on a specific dataset does stay away from as good as another site (B. Agarwal, N. Mittal, 2016 and G.Yoo, J. Nam, 2018).

2.2.4.2.1. Reinforcement Learning Technique

In all, the techniques show how to make the best decisions, an important technique that is relatively different from the unsupervised counterpart. This technique aims to improve text classification's efficiency to show that reinforcement learning techniques are essential and prominent. There are three approaches to implementing a Reinforcement Learning algorithm.

Value-Based: In a value-based Reinforcement Learning method, you should try to maximize a value function $V(s)$. In this method, the agent expects a long-term return of the current states under policy π .

Policy-based: In a policy-based R.L. method, you try to devise such a policy that the action performed in every state helps you gain maximum future reward.

Two types of policy-based methods are:

Deterministic: For any state, the same action is produced by the policy π .

Stochastic: Every action has a certain probability, determined by the following equation. Stochastic Policy:

$$N \{ a \setminus s \} = P \setminus A, = a \setminus S, = S$$

Model-Based: In this Reinforcement Learning method, you must create a virtual model for each environment. The agent learns to perform in that specific environment.

2.2.4.2.2. Unsupervised Approach

This unique machine learning algorithm is often used to make various inferences about data. These datasets consist of labeled, unresponsive input data. Use when labeled training material is not available.

Many current tactics for sentiment examination depend on monitored understanding types trained by labeled corpora, each document becoming manufacturer before education (Ruge *et al*, 2012). However, sometimes, that is complex to accumulate and produce visible datasets (Kalal *et al*, 2019), specifically suited to textual data, that may be undoubtedly unstructured the majority related to the period. Will certainly the fault their very own era requires men and women to label data which in change is actually labor-intensive and labor-intensive.

About usually, is better to accumulate unlabeled datasets and then classify those utilizing unsupervised learning strategies. These techniques take advantage of the documents' report properties, such while phrase co-occurrence and NLP techniques, and present lexicons with psychological (or) polarized keywords (H. Sankar, V. Subramaniaswamy, 2017). Nevertheless, within just system learning, unsupervised approaches throughout the field, including sentiment research, usually use clustering, which can sort documents into different types without indicating exactly which belief will probably be symbolized by every course. In other words, the clustering strategy splits files into groupings (clusters), in which group files are similar with a specific level including Check out to the

data involving different groups. (B. Ma, H. Yuan, Y. Wu, 2017) the performance involves suitable standard clustering methods with admiration for the emotion evaluation. In compliance with team research methods can be arranged straight into Hierarchical and partition clustering.

Hierarchical Method: Hierarchical methods create some kind of hierarchical decomposition, including a dataset displayed by nested organizations (groups that might have sub-groups) ready as a woodland. Hierarchical techniques can frequently be separated into two essential methods, especially Agglomerative and Divisive clustering (H. Suresh, S. Gladston Raj, 2017). Generally, Divisive clustering is honored as the top-down strategy. This specific approach starts with your platoon that groups all the details and also designates those data into sub-clusters using a recursive process with good parallels. (Tsagkalidou *et al.*, 2011), the strategy used this technique to suggest a clustering program that golfing clubs' websites have good nearness they show certain feelings. Usually, The particular Agglomerative clustering (likewise the bottom-up approach) appears that every single data starts inside its quantum and includes the atelier that might have equivalent data until a single or other interactions stay.

Partition method: The partition method can partition data into fixed, non-overlapping groups where every simple factor is given to one specific cluster (Cui *et al.*, 2005). This partitioning depends on similarity requirements which generally could be the Euclidean distance including elements. The documents in a group include a rapid distance to another while typically using the most substantial distance to the information regarding other clusters.

2.2.4.2.3. Supervised Learning

This particular model literacy type utilizes a training dataset to make prognostications. Death records contain both input data and response ideals. Supervised literacy styles use several different training documents. Viewed methods bear designated training documents when the markers are usually the assignments (e.g., positive, neutral, and negative). For illustration, vibrant, practically viewed order approaches might be direct, probabilistic, rule-grounded, and selection wood (H. Sankar, Sixth is v. Subramaniaswamy, 2017). The nicely easy description and a great analysis of usually the most monitored academy approach constantly employed regarding emotion exploration.

2.2.4.2.4. Probabilistic Classifiers

Numerous designs in probabilistic divisors are employed for conferences. There are many types of blended models. Each crossbreed model must end up being an intertwined chemical substance element. Each sort of this paste has a generative effect and could support each getting pregnant by adding this specific aspect or additional factors. This method is named a technology classifier.

i. Maximum Entropy Classifier: The maximum entropy classifier is a classification commonly used in NLP, language, data, and addressing issues. Maximum entropy is also an estimate of the probability distribution. This is an important and well-known technique widely used for various natural language tasks, including language modeling, part-of-speech tagging, and text segmentation. The underlying principle of maximum entropy is the lack of external knowledge.

ii. Bayesian Network classifier: Typically the critical premise of any Bayesian network series is established of variables, every variable containing a new finite pair of communautaire cases. It doesn't depend on the functions.

Typically, the initial supposition is that an offer may have each package dependent on that. This guided chart contributes to a specific Bayesian network type representing arbitrary deals.

iii. Naïve Bayes: Naive Bayes is the most popular textbook bracket system currently. The Naive Bayes bracket model calculates the backward chances of a class grounded on the word splits in the accepted document.

2.2.4.2.5. Rule-based Classification

Rule-grounded groups are used in schemas that make groups according to IF and also rules.

Utmost of the meaning of rule- grounded bracket helps you consider any bracket scheme that uses IF- also rules for class vaticinator (A.K.H Tung, 2009). Thus, the classifiers linked to it count on several guidelines to perform feelings brackets. LHS can explain several feathers of principle

2.2.4.2.6. Linear Classifier

This is a decision based on the value of the linear combination of features. Object properties, also known as feature values, are usually presented to the machine in feature vectors. Linear classifiers can be divided into two methods. They are:

i. Neural Network: Neural networks are a series of algorithms based on recognizing relationships that are unique to multiple datasets, using a process similar to how the human mind works.

ii. Support vector machine: SVMs are used to analyze datasets for classification and regression analysis. This is a machine learning algorithm for processing data automatically.

2.2.4.2.7. Decision Tree Classifiers (DTC)

Decision Tree Classifiers are used for classification. Its purpose is to divide extensive data into smaller groups for easy control. DTC uses multiple values of data attributes and characteristics to get individual predictions for class labels. This is a straightforward technique that is widely used in the field of sentiment analysis.

Utilizing such kind of strategy, education information space is deconstructed hierarchically, utilizing the situation on the particular attribute value to categorize suggestions data into a finite quantity of pre-defined courses. The situation on attribute values will be the existence or absence of numerous words (Medhat *et al*, 2014). This dependent hardwoods strategy is the flowchart-like framework, where each inner customer denotes a check with and perform, each branch signifies results associated with the test, and tea leaf systems symbolize kid systems or course Droit (Han *et al*, 2012). Choice wood dividers are simple to appreciate and perhaps convert; additionally, they can cope with noisy data. However, this type of person is unpredictable and susceptible to over-fitting (Nisbet *et al*, 2018). Furthermore, the decision woods strategy performs completely on large datasets; hence it will probably not be recommended considering small datasets.

2.2.4.3. Hybrid Techniques Approach

This mongrel strategy includes multiple computational processes that provide much more benefits than private approaches and increase emotional (data) research. This fashion advantages numerous, including two and other technologies with significantly better effects than the number of models.

The majority of the hybrid approach combines machine learning and even lexicon-based methods. Get across types is some type of name that makes reference to the blend of machine mastering and lexicon-based methods for feeling examination. The particular cross forms of technology combine both and are absolutely popular, with experiencing lexicons actively playing some sort of component throughout almost all devices. Sensation analysis can be a hybrid strategy, like both record and even knowledge-based methods intended for polarity recognition. Throughout the work including Hassonah *et al*, (2020a) proposed some sort of cross-machine mastering strategy using SVM and two characteristic selection techniques employing the multi-verse windows optimizer

2.2.4.4. Deep Learning Approach

Producing using ANNs-based serious mastering (DL) to feel analysis features has become really favored recently. DL is certainly an aufstrebend host to equipment mastering that materials choices for perfecting efficiency rendering during a supervised or perhaps unsupervised fashion (Rojas *et al*, 2016). Subject “deep learning” the ability to be able to detectors organs internet sites with multiple divisions of perceptions prompted by our brain (Vateekul *et al*, 2016). On that basis, it is potential in this structure to be able to have the ability to be able to teach more revolutionary types than a more excellent dataset and, therefore, produce advanced strengths in many apple iPhone application domain names, including computer system vision and even presentation acknowledgment to be able to NLP (Zhang *et al*, 2018).

Deep learning includes a lot of nerve organ community models such kind of while CNN (Convolutional Nerve organs Networks) (Kim, 2014), RNN (Recurrent Nerve bodily organs Networks) (Li *et al*, 2020), and DBN (Deep Belief Networks) (Zhou *et al*, 2014). These models carried out their most certainly should not find pre-defined features handpicked by a professional engineer. However, they can analyze intricate features throughout the dataset (Shirani-mehr, 2015). On most of the other side, they may be challenging and even computationally very high-priced. Various studies discussed serious learning strategies intended for sentiment examination through detail (Dang *et al*, 2020), (Sohangir *et al*, 2018). Even so, the following subsections give brief information and a summarization concerning most of the most repeated serious mastering models employed for experience analysis.

Deep neural networks (DNN): That is undoubtedly a Man-made Nerve organs Group (ANN) using multiple tiers (hidden layers) between typically the output and type layers (Schmidhuber, 2015). What sort of sub-caste involves type data? Generally, the hidden layers blend control bumps known as neurons. The outgrowth sub-caste influences one or more colorful neurons to deliver the city results (Para *et al*, 2020). That utilizes complex statistical building and the literacy electric power involving ANN to get the real love, whether primary or non-linear, to collude the type into the matter. The stable flush fashion of ANNs and DNNs includes feed forward and backward. Feed forward ANNs will be specific sites and even, so they operate for the feeling category. DNN buildings and their particular alternatives have been employed in many NLP tasks, including feeling examination. (Vassilev, 2019) produced a style known as BowTie based on a severe feed-forward nerve organ community; it requires one signal part, a cascade involving invisible layers, and even an output part. Generally, the analysis of this unit shows charming benefits when compared to suitable other styles.

Convolutional Neural Network (CNN): This unique structure is generally some type regarding the particular type of dental appliance concerning oral appliance concerning feed-forward neural program at first used in the spot

concerning computer eye-sight (Ouyang *et al*, 2015). However, it has just lately accomplished prosperous positive aspects within numerous areas, such as recommender techniques, in addition to NLPs. The quantity involving a CNN incorporates a recommendations part and a finish result part, along with a new invisible coating that will consist of multiple convolutional divisions, pooling levels, and normalization layers in inclusion to becoming a member of levels. Convolutional levels of filtration typically the particular advice (e.g., phrase embedding all through text message emotion classification) to draw out capabilities, although gathering tiers lower your picture resolution regarding functions to create function detection third celebration of sound plus even tiny modifications.

Recurrent Neural Network (RNN): This kind uses a storage area space cell to be able to procedure an integral part of tips. Typically the likelihood to capture and keep advice about the long sequence will make RNNs extensively utilized for NLP jobs like belief analysis (Aziz *et al*, 2018). Within RNNs, the effect typically relies mainly on, after all, the before computations. Regarding occasion, to anticipate the particular subsequent phrase inside a new sentence, the design, and style utilizing the previous words' claims and the relationship between them (Chen *et al*, 2019). One of typically the fundamental problems regarding regular RNN is usually disappearing gradient. Also, to conquer this trouble, Hochreiter and Schmidhuber (Hochreiter *et al*, 1997) created a unique sort linked with RNN referred to as Long-Short Expression Storage (LSTM) that may become popular in numerous career fields. These specific buildings remain significantly utilized by several scientists with consideration to sentiment category. (Li *et al*, 2020), Recommended online LSTM style which can exploit the textual content among target terms plus sentiment polarity phrases in a new phrase without counting after any belief lexicon. The particular trial and mistake outcomes show that this type outperformed other superior methods.

2.2.5. Output Presentation

The main goal of assaying big data is to transfigure an unformed book into helpful tips and screen it in charts comparable to charts, line charts, and bar charts.

2.3. Detection of Emotion

Ideas are an essential part of the mortal lifestyle. These feelings impact human decision-making, in addition, to helping facilitate connection with the surroundings. Emotion recognition, generally known as emotion recognition, is owned by a person's feelings (joy, hopelessness, wrathfulness, etc.). Experimenters have qualified to automate experiencing acknowledgment in the latest times. Still, several physical conditioning, such as heart rate, hand tremors, excessive perspiration, and pitch, express a new person's mental state (Kratzwald *et al*, 2018). Nevertheless, feelings identification from the textbook is usually veritably delicate.

Furthermore, a vast array regarding inscrutability and brand new shoptalk and lingo introduced daily can make it very soft to tell component emotions from this textual content. Also, experiencing acknowledgment isn't nominal regarding important mental countries(pleasure, misery, wrathfulness). Alternatively, chances are to attain upward of six or eight weighing scales structured on the internal model.

2.3.1. Emotion Theories/ Emotion Models

The word "emotion" comes from the 17th century, and the French word "emotion" means disability. Before the 19th century, passion, appetite, and tendencies were categorized as mental states. In the 19th century, "emotion"

was considered a psychological term (Dixon, 2012). In psychology, complex emotional states lead to changes in thoughts, behaviors, behaviors, and personalities called emotions. Psychological or emotional models generally fall into two categories: dimensions and categories.

Emotion Dimensional model: This expresses feelings grounded on three guidelines valence, thrill, plus force (Bakkeretal., 2014). Feelings imply opposition and arising means how instigative the feelings are usually. For instance, pleasure is more instigative than happiness. Energy or domination indicates restraint on emotions. These parameters figure out the position associated with the cerebral condition in two-dimensional space, as demonstrated in Figure 2.3

Emotion Categorical model: The particular categorical model jointly defines feelings much like wrathfulness, happiness, unhappiness, and fear. Based on the older model, emotions fall into four, six, or eight orders.

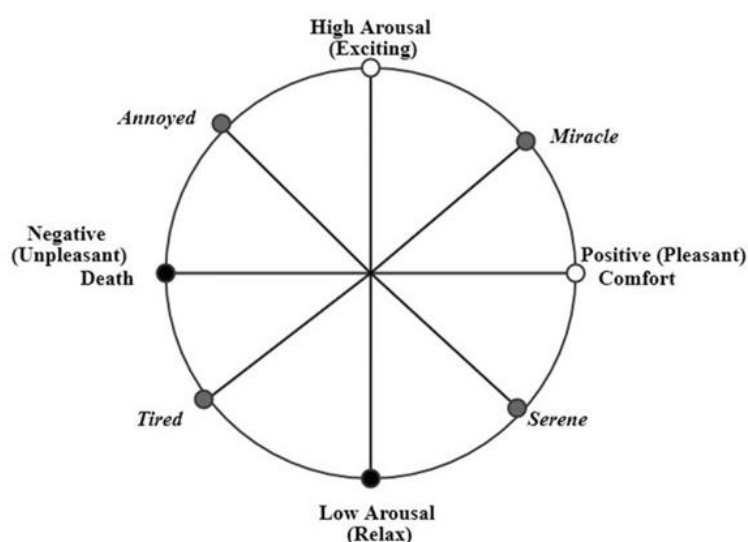


Figure 2.3. Dimensional model of emotions (Alessia *et al.*, 2015)

2.4. Pre-processing of Text

On social media, people usually convey their feelings and feelings effortlessly. As a result, the data collected from blog posts, audits, feedback, views, and criticisms on this social media platform is much unstructured, making it difficult for machines to analyze emotions and emotions. Therefore, pre-processing is an essential phase of data cleansing, as data quality significantly impacts many post-processing approaches. In addition, organizing a dataset requires pre-processing, such as tokenization, stop-word removal, and part-of-speech tagging.

(Abdi *et al.*, 2019); (Bhaskar *et al.*, 2015). Some of these pretreatment techniques can result in losing important information for mood and emotion analysis and must be addressed.

Tokenization is breaking an entire document or paragraph or just a sentence into blocks of words called tokens (Nagaraja *et al.*, 2019). For example, consider the sentence "this place is wonderful," and after tokenization, it will be "this," "place," "yes," good, standardize the text, correct the spelling of the word, etc.

Unnecessary words such as articles and prepositions that do not contribute to emotion recognition or sentiment analysis should be removed. For example, stop words such as "is," "at," "an" and "the" have nothing to do with

emotions and should permanently be removed to steer clear of unnecessarily calculations. Part-of-speech tagging is a way to identify different parts of speech in a sentence. This step helps find different aspects of a sentence commonly represented by a noun or noun phrase, but thoughts are presented adjectives (Sun *et al.*, 2017).

Stemming and lemmatization are important ways of pre-processing. In stemming, expressions have been converted to their root shape via abridging suffixes. For illustration, the terms "argued" and "claim" come "discuss." This procedure reduces the undesirable calculation of rulings (Kratzwald *et al.*, 2018); (Akilandeswari *et al.*, 2018). Lemmatization includes morphological evaluation to exclude inflectional consummations from a commemorative to show it into the bottom expression lemma (Ghanbari *et al.*, 2019). For case, the term "caught" is converted into "catch" (Ahuja *et al.*, 2019). (Symeonidis *et al.*, 2018) tested the overall performance of 4gadgetsstudyingfashions with a total and ablation to look at different-processing strategies on datasets, specifically SS-Tweet and SemEval. The authors concluded that putting off figures and lemmatization is more suitable for delicacy while putting off punctuation no longer affects fineness.

2.5. Application of Sentiment Analysis

Idea analysis is very within many plan fields starting approaching from identifying consumer view (Roy *et al.*, 2019), (Bose *et al.*, 2020) to be able to supervisory mental well-being dedicated to patient's social media marketing and advertising content (A. Rajput, 2019). In addition, to manage this, most of the beginning of new-technology for example, Huge Info (Yaqoob, *et al.* 2016), Fog up Computer (S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, A. Ghalsasi, 2011) as well as Blockchain (Frizzo-Barker *et al.*, 2020) has increased almost all the area regarding programs providing opinion analysis together with endless possibilities to be able to get utilized inside nearly every one website. For occasion, several of almost all of the frequent app websites regarding feeling research usually are referred to in the seeking subsections.

Business Intelligent: Generating using emotion research in the domain of business intelligence contains several positive aspects; by way of example, companies could get good things about typically the outcomes regarding belief evaluation for generating products or services improvements, look at most of the customer's ideas, or adopt a new manufacturer-new marketing and advertising approach (Bernabé-Moreno *et al.*, 2020). Inspecting customers' awareness involving items or companies is typically the most recurrent application of sentiment exploration, regarded as a new way of enterprise intelligence. Alternatively, these kinds of analyses typically are generally not suitable only for product or service manufacturers. Nevertheless, customers can help to be able to manage to make use of just about all involving them and to look at companies make a new far better decision.

Recommendation system: The existing recommender system will usually be produced and will end up being built to recommend related items (movies, tracks, or items inside order so as to buy) to clients (Z.Y. Khan, Z. Niu, S. Sandiwarno, R. Prince, 2020). An excellent successful recommender program could produce a lot in earnings for several industries. Therefore, this specific kind of method (J. Serrano-Guerrero, J.A. Olivas, F.P. Romero, 2020 B. Ray, A. Garain, R. Sarkar 2021, X. Fu, T. Ouyang, Z. Yang, S. Liu, 2020) may enjoy the particular application of emotion analysis typically to create new much better advice. Within the specific function of (Li *et al.*, 2016). The freelance writers suggested KBridge, a brilliant new video suggestion program applying emotion evaluation regarding micro blogs.

Healthcare and Medical System: Applying emotion analysis inside the medical website has lately captivated great interest. This specific function allows health-related actors to gain information regarding esteem, adverse drug replies, pandemics, and cases' moods (Ramírez- Tinoco *et al*, 2019). It examines almost all of them to supply much better health care services. Nevertheless, this is hard in order to make use of feeling evaluation in this type of website as a result of this associated with many encounter issues such as vocabulary due to the fact seen in the particular activity (Zafra *et al.*, 2019). (Clark *et al.*, 2018) decided and examined Twitter posts related to being able to the particular consumer experience due to yet another helpful tool for looking at open public wellness.

Authorities Cleverness: And in addition to products and organizations, folks also develop a commentary on multitudinous subjects, like country-wide politics, opinion, and social concerns. Generating opinion analysis to tell part thoughts about government plans or even similar issues will be highly significant for checking achievable community responses towards executing specific guidelines within the effort (Georgiadou *et al.*, 2020).

2.6. The Challenges with Sentiment Analysis

Difficulties related to emotion analysis are generally tools meant to around the journey of the coaching model. Commentary along with neutrality or natural tone tends in order to beget problems along with the system and it is frequently unknown. With regard to instance, a customer receives a product in the wrong color, and commentary, "The item has been blue," will be linked as natural in order to should become negative.

Also, in case the system doesn't understand the atmosphere or tone, this can be intense to identify the particular mood. For example, answers to inspections and check queries similar as "None" and "All" may be labeled appreciatively or negatively based on the question, which makes it delicate to sort out without a particular environment.. also, degradation and affront are not able to be easily qualified and often results in mislabeled sentiment.

Personal computer programs also possess problems when these people encounter emojis or even inapplicable information. Therefore, particular attention ought to be paid in order to educate strategy along with emojis and natural data to prevent mislabeling the book. In the end, people may be inconsistent with their statements. With regard to instance, utmost evaluations contain both good and negative comments. This is often nicely handled by assaying the particular rulings one simply by one. Still, the greater informal the press, the more probably people will mix different opinions within the same view, making it more difficult for computers in order to dissect them.

2.7. Review of Related Work

Recently, numerous experimenters have tried to incorporate the generalities of deep literacy and machine literacy for sentiment analysis. This kind of section briefly details the multitudinous studies related to feeling analysis optimization with emotional view alerts in machine literacy ways.

Binali *et al.* (2009) proposed an emotion recognition system in e-learning. The authors specified that the system possesses the ability to classify student opinions about learning progress. The authors utilize Gate software to implement the framework. Their implementation uses features like smile, fear, anger, happiness, and sadness in analyzing the system. The result shows better and more flexible performance when considering sentiment analysis, as claimed by the authors.

Chan *et al.* (2012) considered working on the development of an emotion classification system using machine learning algorithms. The authors utilize support vector machines (SVMs) and other algorithms to classify emotional signals from patient datasets. The authors claimed that the techniques used so far projected that SVMs have achieved the highest accuracy, improving system performance by experimenting with different combinations of emotional signals.

Xia *et al.* (2013) proposed an unsupervised sentiment analysis with emotional signals. The authors incorporate signals into an unsupervised learning framework for sentiment analysis. The authors explore a unified way to model two significant categories of sentiment signals: emotion indication and emotion correlation (Xia *et al.*, 2013). The authors further compare the proposed framework with the latest methods of the two Twitter datasets and empirically evaluate the framework to gain a deeper understanding of the effects of emotional signals. Their study used two publicly available tweet datasets: Stanford Twitter Sentiment and the Obama-McCain Debate. The result obtained in their research compared to *GI-Label* has achieved about 21.40% and 17.87% improvement on the two datasets, respectively. Furthermore, the author specified that ESSA-opt's performance is better than ESSA's.

Socher *et al.* (2013) proposed a deep learning module for finely classifying sentences on the corpus of tree banks. The authors utilize recurrent neural network modules, which are designed using training and test datasets to achieve more excellent performance than existing ones. The result of the deep learning module shows that the performance of the proposed techniques amounts to 80-85% accuracy which is achieved compared to the baseline method, as claimed by the authors.

The proposed work (Li *et al.*, 2014) builds a tree bank of Chinese views on social data to overcome the lack of a large corpus labeled in existing models. The authors stated clearly that when predicting labels at the statement level, i.e., positive or negative, the recursive neural deep models (RNDMs) have been suggested, which achieve higher performance than SVM, Nave Bayes, and Maximum Entropy, respectively. The authors reviewed about 2270 movies collected from the site, and these reviews were segmented using the Chinese word segmentation tool ICTCLAS. Five classes were specified for each sentence, and the Stand ford parser was used to analyze the sentence. The results show that the model improved the prediction of the emotional label of the sentence by completing 13550 Chinese sentences and 14964 words. Furthermore, the authors stated that M.E. and N.B. perform better than baseline with a large margin due to the contrasting coupling structure.

Gaurav *et al.* (2014) considered a survey of various machine learning methods for text classification. Their goal was to compare the effectiveness of applying machine learning techniques to the sentiment classification problem. Guarav et al. (2014) introduced various types of engineers for text classification, which provides theoretical and empirical evidence that SVM is more suitable for text classification than other classifiers. The authors claimed that the analysis allows SVM to have higher accuracy and automatically searches and customizes parameter settings. Their research utilizes three different algorithms, namely: Naive Bayes, Support Vector Machines, and Decision Trees using pre-defined data. Linear SVM was the most accurate method with an accuracy of 91.3% in the ten most common categories and 85.5% in all 118 categories respectively. The authors claimed that using relevant results and examples, their work proves that SVM is one of the better algorithms because it provides higher accuracy than the (Naive Bayes and Decision Tree) algorithm.

An automatic feedback analysis of student feedback was proposed by Kaewyong *et al.* (2015) Using a lexicon-based method. The authors used a primary source of data techniques by the help of administering a questionnaire to over 1100 students in responses to education. After applying various pretreatment techniques, their implementation comprises opinion words assigned by an emotion score using an emotion lexicon. The authors claimed that using an emotion lexicon improved and increased the performance of sentiment analysis optimization compared to the primary method.

The author (Severyn & Moschitti, 2015) proposed a deep learning system for Twitter sentiment analysis. Its main goal is initializing the parameter weights of the convolutional neural network, and it is essential to train the model accurately without adding new features. The author used neural language to initialize word embedding, which was trained through a large set of unsupervised tweets. In addition, the author uses components such as activation, sentence matrix pooling, softmax, convolution layer, training the network using Stochastic Gradient Descent (SGD), and a non-convex function optimization algorithm. Finally, the authors applied deep learning to two tasks proposed by Semeval 2015, which include; message-level tasks and phrase-level tasks, to predict polarity and achieve high and better results performance, and the model is ranked first in terms of accuracy as claimed by the authors.

A detailed survey by (Yanmei & Yuda, 2015) provides an overview of sentiment analysis related to microblogging. The main goal was to use a convolutional neural network (CNN) to understand user opinions and attitudes about hot events. Input URLs and focused crawlers were used to collect data from the target, and 1000 microblogging comments were collected as a corpus and split into three labels, i.e., 274 neutral emotions, 300 negative emotions, and 426 positive emotions for their research. The authors use algorithms like CRF, SVM, and additional traditional algorithms to perform sentiment analysis at a high cost. However, their performance proves that the model (Yanmei & Yuda, 2015) is reasonable and sufficient to improve accuracy in sentiment analysis.

Sun *et al.* (2016) proposed a cognitive model for interpreting emotions from complex texts. The authors' analysis of four modules comprises non-behavior-oriented, metacognition, behavior-oriented, and motivation. First, the authors extracted emotions from various tweets using the emotion-word hashtag and the Hashtag Emotion Corpus dataset (Mohammad & Kiritchenko, 2015). In the process, a comprehensive word emotion dictionary was created using the emotion-tagged tweet dataset. The results from the four modules show that the SVM classifier performs better for basic emotion types, as claimed by the authors. However, the research did not consider emotional words with different synonyms, which, when inculcated, can improve the system's performance.

Winarsih and Supriyanto (2016) evaluated the performance of various machine learning classifiers such as Artificial Neural Network, Support Vector Machine, and Naive Bayes, as well as the minimum optimization of emotion classification from Indonesian texts. Furthermore, the authors applied various pre-processing steps such as tokenization, stop-word removal, and case sensitivity. The research experiments are performed with 10-fold cross-validation, demonstrating that Minimization Optimization Technology (SVMSMO) is superior to the comparison method, as claimed by the author. It was finally observed that the result from Winarsih and Supriyanto (2016) possesses a high level of acceptance when considering using machine learning classifiers in sentiment analysis.

Baecchi *et al.* (2016) proposed a model that analyzes the mind based on the Text and visual content of social networks. The authors observed that neural network-based models are superior to N-gram models. The authors utilize the consideration of multimedia techniques to address microblogging dynamic analysis using neural networks. The motivation behind their technique is training in a neural network-based language paradigm on its effectiveness in processing texts on the Web, especially the grammar and semantic similarities of words with uncensored learning. The authors trained four datasets, including Sanders Corpus, Sentiment140, SemEval2013, and SentiBank Twitter. The results show that their model outperformed CBOWS + SVM and FSLM (Full Supervised Prospect Language Model) such that it achieved the best performance and saw the need to learn and conquer large amounts of data, as claimed by Baecchi *et al.* (2016).

Ibrahim *et al.* (2016) developed a sentiment analysis model for people's opinions and feelings about some of the comments collected on Facebook. The authors utilize a secondary source of data techniques, such as comments collected from Facebook, written in Arabic in both task and intensity (Ibrahim *et al.*, 2016). Their main goal was to analyze the impact of pre-processing operations such as noise rejection, stemming, and normalization of user comments. The author claimed that their implementation had received a commendable efficiency, increasing accuracy and system performance in sentiment analysis and optimization.

Jiang and Qi (2016) proposed China's emotion recognition system for classifying user emotions from online product reviews. The authors use the extended OCCOR emotion model in their research work by selecting six emotion categories as its features. The models were evaluated using a variety of machine learning and natural learning techniques as claimed by the authors. The results show the effectiveness and excellent performance of the system regarding sentiment analysis.

Vateekul *et al.* (2016) proposed two deep learning methods for emotion classification of Thai Twitter data: convolution neural networks (DCNN) and short-term memory (LSTM). The authors collected their data from Thai Twitter users and followers. After filtering the data, only users of Thai tweets and tweets containing Thai characters were selected for the experiment. Then, the authors conducted five experiments to achieve optimal parameters for deep learning, compare deep learning with classical methods, and achieve the importance of word order. The results show that DNN is more accurate than LSTM and that both deep learning techniques are more accurate than SVM and Nave Bayes but less than the maximum entropy claimed by Vateekul *et al.* (2016).

Singh *et al.* (2017) proposed using a machine learning classifier to optimize sentiment analysis. The authors' main goal is to extract both positive and negative polarities from social media text, making a sentiment analysis task in natural language processing. For optimizing sentiment analysis, the authors considered four state-of-the-art machine learning classifiers, which comprise Naive Bayes, J48, BFTree, and OneR. That dataset is obtained from Amazon and IMDB Movie Review, respectively. The result from the experiment shows that Naive Bayes has proven to be very fast to learn. However, OneR seems more promising in producing an accuracy of precisely 91.3%, F-measure 97%, and correctly classified instances 92.34%, respectively.

Umar *et al.* (2017) have developed a comprehensive educational model for detecting tweet polarization in five classification levels, from very positive to very negative. The authors optimized tweets from 12 Arab countries in

four different regions, which are comprised (of North Africa, the Arabian Gulf, Egypt, and eastern Arab) respectively. They collected about 470,000 of its twins. Umar *et al.* (2017) used a deep learning model consisting of a content layer followed by an LSTM layer. The pre-trained word theme in Urma *et al.* was applied using Word2Vec's Skip-gram template, Egyptian wedding tweets, and parties which became the data source for Umar *et al.* (2017). It was discovered that the system achieved about 70% accuracy in its implementation, as claimed by the authors.

Maha *et al.* (2018) proposed a split and conquest methodology that performs Sentiment Analysis individually for each type of sentence. The author discovered that sentences tended to be very complicated, mainly when they contained many sentimental words. Therefore, their work proposed to use an NN-centric sequence model to classify selfish statements into three types, depending on the number of targets that occur in the sentence (Maha *et al.*, 2018). Each pool of sentences was then delivered individually to a one-dimensional CNN for sentimental classification. Their approach was evaluated using four sentimental classification datasets and compared to a broad baseline. The authors' results show that: categorizing sentence types improve Sentiment Analysis performance at the sentence level. Furthermore, the model used outperforms the latest F1 deep learning model by 53.6%, achieving 65% accuracy and 64.46% F1 scores claimed by the author.

Hassonah *et al.* (2019) explored an efficient hybrid filter and evolutionary wrapper approach for sentiment analysis on various Twitter topics. The authors utilize a hybrid machine learning approach to improve sentiment analysis by using a Support Vector Machine (SVM) classifier to create a classification model based on three classes, namely: positive, neutral, and negative emotion, combining two feature selection techniques using the Relief F and Multi-Verse Optimizer (MVO) algorithms. The research also extracts over 6900 tweets from Twitter social networks to test the validation of the results. The results show that the Hassonah *et al.* (2019) method outperforms other methods and classifiers by achieving better results on most datasets while reducing the number of features by up to 96.85% compared to the original feature set, which is seen to be excellent.

Ray *et al.* (2020) created an ensemble-based vacation resort recommender system using sentiment evaluation plus aspect categorization associated with hotel reviews. Almost all researchers employ fresh, rich, and various datasets concerning online hotel thoughts indexed from Tripadvisor. Which applied some specialized technique that very first makes using typically the ensemble involving several sorts of the binary class named Bidirectional Régler Illustrations from Transformer remanufacture (BERT) type by using a new few levels regarding positive-negative, neutral-negative, neutral-positive comments combined by simply by using an excess weight assigning protocol. The dataset obtained grouped the reviews into different categories using an approach that involves fuzzy logic and cosine similarity. The researcher prepared the datasets based on crawling data using the Trip advisor API. The crawled dataset consists of 58 612 reviews. The proposed model has achieved a Macro F1-score of 84% and test accuracy of 92.36% in classifying sentiment polarities. The results are pretty promising and much better compared to state-of-the-art models.

Naseem *et al.* (2020) presented COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. The researcher aims to identify the topics and the community sentiment dynamics expressed on Twitter about COVID-19. The authors also analyze views concerning COVID-19 by focusing on people who

interact and share social media on Twitter. They use a new large-scale sentiment data set, COVIDSENTI, which consists of 90 000 COVID-19-related tweets collected using the Twitter API in the early stages of the pandemic, from February to March 2020, and sentiment identified from the collected tweets were labeled into three categories: those containing positive, negative, and neutral tweets. The time-aware knowledge extraction (TAKE) methodology was employed in the research. The study revealed that Negative opinion played an essential role in conditioning public sentiment; for instance, it was observed that people favored lockdown earlier in the pandemic; however, as expected, sentiment shifted by mid-March. Therefore, the study supports the view that there is a need to develop a proactive and agile public health presence to combat the spread of negative sentiment on social media following a pandemic.

Nagamanjula and Pethalakshmi (2020) proposed a new framework based on multi-objective optimization and LAN2FIS for sentiment analysis on Twitter. Twitter topics are so diverse that it is difficult to collect data in emotion classification (Nagamanjula & Pethalakshmi, 2020). Therefore, the authors utilize a new framework for pre-processing information to enrich tweets. As the tweet is processed, various features are extracted from the tweet. To take advantage of this vast amount of information, a proposed hybrid machine learning algorithm called LAN2FIS (Logistic Adaptive Network Based on Neurophagy Inference System) (Nagamanjula & Pethalakshmi, 2020). Their work presents objective biological optimization (minimum redundancy and maximum association) for feature selection and finds that more efficient feature subsets can be obtained. The result evaluates performance in terms of accuracy, precision, recall, F-Measure, and error rate: displaying that HMLA is more efficient and accurate than other classifiers.

Muhammad *et al.* (2020) research on performance analysis of supervised machine learning techniques for efficiently detecting thoughts from online content. The authors claimed that most of the existing work on emotion detection suffered from poor performance due to inefficient machine learning classifiers with limited datasets (Muhammad *et al.*, 2020). To solve such a problem, the authors' goals aim to evaluate the performance of various machine learning classifiers on benchmark sentiment datasets. The authors trained their proposed method with machine teaching classifiers such as (Random Forest, SVM, Logistic Regression, Xgboost, SGD Classifier, Naive Bayes Classifier, and ANN). Muhammad *et al.* (2020) claimed that the experimental results on precision, recall, and f-Measure show that the logistic regression classifier outperforms other classifiers in terms of improved recall with an accuracy of (83%), with BPN achieving improved accuracy of (71.27%). In contrast to the result achieved, the SVM results achieved an accuracy of (76%) and the f-score achieved an accuracy of (77%) respectively. The authors claimed that at worst-case analysis, XGBoost shows poor performance in terms of reduced accuracy by (66%), recall (66%), F-measure (66%), and accuracy (58.5%), respectively.

Strimaitis *et al.* (2021) Delved rewarding environment mass media sentiment analysis regarding Lithuanian Language. Usually, the exploration's main mendicite is datasets regarding sentiment analysis regarding fiscal news, along with looking for just how effective being mount styles are regarding feeling estimation together with this dataset. The particular work used the particular positive and bad bigrams prepared simply by monetary company professionals to perform the particular wordbook-grounded approach. The particular query utilizes the particularly closely watched device literacy model, which has been forced to find out the stylish series for its gathered dataset.

The lately accumulated dataset is usually obtained from 4 leading Lithuanian reports websites, which had been accumulated and experimentally anatomized (Lithuanian Economic Media 2021). The particular examination was executed using three typically used bracket methods in the industry of sentiment research the multinomial Naïve Bayes, an help vector machine, addition to prolonged short-term storage conditions were anatomized and compared. The final results of the used machine literacy procedures show that the particular loftiest delicacy is usually attained utilizing an on-balanced dataset with the multinomial Trusting Bayes algorithm (71.1%). Another conditions rigor was clearly a little below an extended immediate memory(71%) in addition to an aid vector machine (70.4%) the multinomial Unsuspecting Bayes criteria, in every circumstance, obtains the particular loftiest model treat.

Basiri *et al.* (2021) Lookup of a company new emulsion- grounded strong literacy type regarding sentiment evaluation regarding COVID- 19 Twitter posts. Knowledge- Grounded Methods. Generally, the experimenters advised a fresh approach grounded on typically the particular emulsion regarding 4 serious literacy, in addition, to being able to one time-honored watched machine literacy sort for feeling analysis of coronavirus-related Facebook posts from 8-10 nations throughout typically the world. The examination was conducted to discover people's standard sentiments (opinions) in 8-10 countries worldwide. Facebook posts from individual's 8-10 places between2020-01-24 as well since to2020-04-21 and GoogleTrends druggies were obtained using coronavirus-related crucial term missions from2020-01-24 to2020-04-23. Most of the pursuit utilized two info resources, videlicet Yahoo Developments, and Facebook info. Bing Developments details utilized inside order to dissect people's curiosity to be able to gain information relating to COVID- 19disease by implementing Yahoo and yahoo quests regarding associated keywords. Their very own query says genuinely does the coronavirus mesmerize the curiosity of men in addition to women from diverse international places all over the world at different durations in varying forces. still, the experimenter items out that may the study neglects items regarding the world COVID- 19 reports and numbers within just the overall idea of other nations around the world throughout the planet because of the limitation.

Fayyouni *et al.* (2021) Semantic Partitioning and Items literacy in Discomfort Analysis. The professionals probe sentiment test in Arabic Twitter updates who own the occurrence linked with Jordanian shoptalk gathering 2000 myspace up-dates written throughout The particular nike pas cher air jordan during the COVID- nineteen crisis. Here, typically the examination proposed a couple of versions to prognosticate the particular level of resistance with the particular accumulated Twitter up-dates by simply invoking Assist Vector Machine(SVM), Naïve Bayes (N.B), J48, Multi-Layer Perceptron (MLP), plus Logistic Retrogression(L.R) classifiers. Several sorts of brand-new datasets were accumulated in the coronavirus complaint (COVID-19) epidemic. The problem demonstrated two editions the standard Persia Language (TAL) unit plus the Semantic Dividing Arabic Language (SPAL) model, to be able to picture the levels associated with the weight of the certainly accumulated tweets merely by invoking various recognized divisors. Typically the delivery and portion associated with multitudinous Arabic functions, corresponding to spoken, writing fashion, grammatical, and emotional functions, have just lately already been applied to dissect and additionally classify the accumulated tweets semantically. Typically the particular study outgrowth exposed an enhancement within the handle upward to 4.22% in the SPAL unit when

compared with the TAL unit as soon as the MLP classifier seemed to be invoked. Also, development in the weighted F- credit history, which gifts a fantastic indicator of which can make typically the majority of typically the particular conflict involving remember and flawlessness, the particular SPAL model towards typically the IGUAL design achieved way upward to 8.70% within typically the J48 repertoires. Typically the partitioning in the accumulated Twitter updates is definitely achieved actually, certainly not automatically, considering that the level of resistance is by side assigned for typically the particularly collected myspace updates, which may well be considered it is a limitation.

Villavicencio *et al.* (2021) considered Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes. The study aimed to analyze sentiments towards COVID-19 vaccines in the Philippines according to positive, neutral, and negative polarities. In line with this, the researchers used all the tweets in the first month of the implementation of the vaccination program. The authors gathered data on the sentiment of Filipinos regarding the Philippines government's efforts using the social networking site Twitter. The researchers started by collecting related tweets, followed by data annotation, data processing through NLP techniques, sentiment classification using the Naïve Bayes classifier algorithm, and performance evaluation by applying the developed model in an unlabeled dataset. Then, the sentiments were annotated and trained using the Naïve Bayes model to classify English and Filipino language tweets into positive, neutral, and negative polarities through the RapidMiner data science software. The results yielded an 81.77% accuracy, which outweighs the accuracy of recent sentiment analysis studies using Twitter data from the Philippines. Based on the study outcome, it can be concluded that the majority, 83%, of the tweets in the Philippines were positive and enthusiastic about the idea of vaccination. In comparison, 9% had neutral and 8% had negative sentiments.

Khan *et al.* (2021) proposed U.S. Based COVID-19 Tweets Sentiment Analysis Using TextBlob and Supervised Machine Learning Algorithms. The research aimed to analyze the critical situation for making better policies for U.S. residents. The researchers proposed a US-based sentiment analysis of the tweets using machine learning and the lexicon analysis approach. The authors made use of a US-based COVID-19 raw Twitter dataset created by the Department of Computer Science, Abbott wrong university of Science & Technology for sentiment analysis which was collected by RStudio software from 30 January 2020 to 10 May 2020, containing 11858 tweets that were labeled corresponding to each tweet using TextBlob, into positive, negative, or neutral and the tweets was further pre-processed. The study employed various supervised ML methods, which were used to address tweet classification challenges based on two feature extraction methods, BoW and TF-IDF. The random forest, gradient boosting machine, extra tree classifier, logistic regression, and support vector machine models categorize beliefs as positive, negative, or neutral for US-based COVID-19 Tweets. The research shows how TF-IDF features can increase the performance of the supervised machine learning models, which gradient boosting machine outperforms the others and achieves high accuracy of 96% when paired with TF-IDF features and validate the approach's effectiveness.

Neogi *et al.*, (2021) Sentiment analysis and classification of Indian farmers' protest using Twitter data. The authors aimed to understand the sentiments of the Indian citizens towards the three acts passed by the government by incorporating NLP techniques. In addition, they also analyzed the polarity and factuality of Twitter data regarding

the demonstrations by extracting twitter data. In the research, data were gathered from the microblogging website Twitter concerning farmers' protests to understand the sentiments that the public shared on an international level. The raw data was collected by us using an open-source python library called tweepy to directly access the Twitter API, which in turn uses private access tokens and consumer keys for authentication purposes. Models were employed to categorize and analyze the sentiments based on a collection of around 20,000 tweets on the protest. The research work was analyzed using Bag of Words and TF-IDF to convert the textual information to numeric weightage in vector format.

Furthermore, they used four classifiers, Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine, for prediction purposes. It was compared that Random Forest had the highest classification accuracy with the best result. The author points out that the research lacked the computational resources to process such massive tweets.

Wang *et al.* (2021) considered Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis. The research proposed the RGWE method based on the sentiment concept to achieve the accurate embedding of sentiment information and provide more precise semantics and sentiment representations for words. First, it found the optimal sentiment concept of words in the Microsoft Concept Graph according to the context of words. Then obtained, the sentiment information of words under optimal sentiment concept from the multi-semantics sentiment intensity lexicon, which was constructed to achieve accurate embedding of sentiment information and provide more accurate semantics and sentiment representation for words. Finally, the authors utilized six available classical public datasets (SemEval, SST1, SST2, IDM, Amazone, and Yelp 2014) that were selected to evaluate the performance of RGWE proposed in Sentiment Analysis tasks. The validity of RGWE is verified by comparing it with the traditional embedding and sentiment embedding methods on typical datasets. Furthermore, RGWE integrates different position features and internal and external sentiment information by averaging Refined-Word2Vec and Refined-GloVe, further improving Sentiment Analysis's accuracy.

Sweidan *et al.* (2021) presented Sentence-Level Aspect-Based Sentiment Analysis for Classifying Adverse Drug Reactions (ADRs) Using Hybrid Ontology-XLNet Transfer Learning. The study aimed to detect aspects and identify the sentiments related to each aspect expressed by users in social data. Thus, the authors investigated the contribution of utilizing the lexicalized ontology to improve the aspect-based sentiment analysis performance through extracting the indirect relationships in user social data. The research further proposed an ontology-XLNet-based aspect sentiment analysis approach for ADRs, which consists of three phases: pre-processing, feature extraction, and Sentiment Classification. First, the datasets used in the experimental work of the research are constructed based on a set of drug reviews and Twitter posts extracted from different resources. Next, the user's opinion about drugs is used to detect and extract unreported drug reactions and classify them according to their social data (reviews, posts). Finally, the XLNet model is utilized to extract the neighboring contextual meaning and concatenate it with each embedding word to produce a more comprehensive context and enhance feature extraction. The research revealed that the approach outperformed other tested state-of-the-art related approaches by improving feature extraction of unstructured social media text and overall sentiment classification accuracy. A significant accuracy of 98% and F-measure of 96.4% is achieved by the proposed ADRs aspect-based sentiment analysis approach.

Pansy and Rupali (2021) reviewed sentiment analysis and emotion recognition reviews from TextText. Its main goal is to outline existing methods for detecting emotions and moods. However, while the dictionary-based approach is adaptable and easy to apply, the corpus-based method is based on rules that work effectively in certain areas (Pansy & Rupali, 2021). As a result, the corpus-based approach is more accurate but cannot be generalized. The deep learning approach outperforms machine learning in situations where the dataset is vast. Recurrent neural networks, especially LSTM models, are widely used for mood and emotion analysis because they cover long-term dependencies and are very good at extracting features. However, RNNs with attention networks work very well. At the same time, it is essential to remember that dictionary-based and machine learning approaches (traditional approaches) are also evolving, with better results. Pre-processing and feature extraction techniques also significantly impact the performance of different approaches to mood and emotion analysis claimed by the authors.

Deng *et al.* (2022) advised Text sentiment examination of the emulsion model grounded on the interesting medium. The experimenters recommend an emulsion model to accomplish high flawlessness in the book feeling analysis when the model combines the characteristics of CNN to prize original information of Text and BiLSTM to award the in-text connection Text and introduces the interesting medium to increase the give attention to words with a solid emotional tendency in the text. The advised system is able to the textbook belief trend analysis with a few the considerable weight value of experience vocabulary because it can be appropriate with this is of individual words and Text. Most of the datasets used were the training datasets from the task that crawled from several social multimedia system spots like Facebook or myspace, WhatsApp, and Facebook, In addition to many others. Typically the model presented a unique medium, which makes it pay further attention to the mental word information in common sense at the point of the beginning process and reduces the effect of people's words that aren't important for the mount. The combo of CNN and the BiLSTM model has achieved improved output in treatment than any other model. Still, this requirement doesn't ameliorate the network's internal formula.

2.8. Summary of Literature

Extensive research on the related studies has shown that various machines and deep learning algorithms have optimized sentiment analysis and emotional opinion signals.

Table 1. Summary of the studies related to sentiment analysis

Author and Year	Model Adopted	Purpose	Data set Used	Results
MahaH., 2019	Convolutional Neural Network (DCNN) and short-term memory	To predict the sentiments of Arabic tweets that use the Arabic Sentiments Tweets	On smaller datasets which consist of 10,000 tweets, distributed among four classes (positive, negative, neutral, and objective).	This model outperforms the state-of-the-art deep learning model's F1 score of 53.6% in the Arabic Emotion Tweet Dataset (ASTD), achieving an F1 score of 64.46%.

Ibrahim R., 2018	Sentiment classification and Machine Learning	To analyze people's opinions and emotions towards	4-GB memory size and 2.50 GHz processing speed. The model also was run and tested using three testbeds or datasets.	The sentiment model behaves better using the light stemmer than using the general one accuracy of 70.1%
Ramya. 2017	Sentiment analysis and Long Short-term memory LSTM	Models on the tweets of both Egypt and the	Retrieved tweets from 12 Arab countries in 4 different regions (North Africa, Arabian Gulf, Egypt, Eastern Arab). They gathered 470,000 of his twins.	The results show that the excellent performance of deep learning models, the importance of the morphological features of Arabic NLP, and the processing of the dialect Arabic yield different results depending on the country from which the tweet was collected. 70% accuracy.
Vateekul & Koomashubha, 2016	Convolutional Neural Networks (DCNN) and short-term memory (LSTM)	Sentiment Analysis on Thai Twitter Data	3,813,173 tweets (33,349 negative tweets and 140,414 positive tweets).	Higher accuracy than SVM and Nave Bayes Less than maximum entropy Original sentence accuracy is higher than mixed sentence 60.8%
Baecchi <i>et al.</i> , 2016	Deep Neural Networks (CBOW-DA-LR)	Visual and textual sentiment analysis	Four (4) datasets: sanders corpus, sentiment 140, seminal-2013, and senti-Bank twitter Dataset.	The CBOWSALR model achieves better classification accuracy than previous models. Accuracy 79.39%
Severyn & Moschitti, 2015	Convolutional Neural Networks (CNN)	Phrase level and message level task	Removal-201	Compared to the official system, it is ranked 1st in the phase level subtasks and 2nd in the message

		Sentiment Analysis		level.
Yanmei & Yuda, 2015	Convolutional Neural Networks (CNN)	Micro-Blog Sentiment Analysis	1000 microblog comments (Hua Qiang).	The proposed model can effectively improve the accuracy of emotional orientation and validation.
Li <i>et al.</i> , (2014)	Recursive Neural Deep Model (RNDM)	Chinese sentiments analysis of social data	2270 movie reviews from websites.	Provides higher performance (90.8%) than baseline with wide margins.

2.9. Knowledge Gap

Related works reviewed. It can be identified that the researchers had tremendously contributed using diverse approaches, including machine learning and deep learning techniques. Algorithms such as the Support Vector Machine, and many other classification algorithms have proven successful, but the accuracy level is not yet efficient enough. Furthermore, other authors have applied deep learning techniques while showing promising results but require much training to overcome false alarms of positive sentiment. Nevertheless, machine learning has proven more suitable for classification problems with an algorithm such as the Support Vector Machine, Random Forest, and others taking the lead in classification problems. Hence, this study proposes to develop hybrid model algorithms, namely the Support Vector Machine, and Linear Regression, while optimizing their result.

3. CONCLUSION

This specific awesome article discussed sentiment examination in addition to associated approaches. The specific primary target of the work is undoubtedly to check on plus even total category approaches with their own advantages plus drawbacks throughout emotion evaluation. To become capable to start, a number of numbers related to emotion analyses have already been discussed, associated with just a simple overview including required procedures of this particular type as info series and functionality variety. Next, techniques involving sentiment categorization devices were organized and also hybridize regarding their very own personal benefits plus cons. Due to the fact related to simplicity as well as outstanding accuracy, carefully viewed machine studying methods are generally the particular commonly used method throughout this particular self-discipline. Category making use of SVM as well as LR algorithms will end up generally used whilst standards against which recently proposed techniques could be opposed. Some associated with the most typical software places are often reviewed after that they will study and investigates the really worth and implications connected with sentiment exam difficulties in sensation assessment. The assessment investigates the passionate partnership between structures associated with emotional opinions plus the particular issues linked in order to the sentiment exam. This particular hybrid reveals domain name dependence, which is required with regard in order to identify sentiment issues. Typically future

features can consist connected with constantly growing the particular assessment area along with further findings. The particular future difficulties demonstrate that emotion assessment is unquestionably still the fairly unexplored subject issue associated with the specific study.

Declarations

Source of Funding

This research work did not receive any grant from funding agencies in the public or not-for-profit sectors.

Competing Interests Statement

The authors declare no competing financial, professional, or personal interests.

Consent for publication

The authors declare that they consented to the publication of this research work.

Authors' Contributions

All authors equally contributed to research and paper drafting.

References

- Abdi A, Shamsuddin SM, Hasan S, Piran J. (2019). Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. *Inf Process Manag.*, 56(4): 1245–1259.
- Agaian, S. and Kolm, P. (2017). Financial Sentiment Analysis Using Machine Learning Techniques. *International Journal of Investment Management and Financial Innovations*, 3: 1–9.
- Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., Gupta, B. (2017). Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computing Science*, 27: 386–392.
- Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., Manicardi, S. (2016). A Comparison between Pre-processing Techniques for Sentiment Analysis in Twitter. https://www.researchgate.net/publication/311615347_A_Comparison_between_Preprocessing_Techniques_for_Sentiment_Analysis_in_Twitter/citation/download.
- Appel, O., Chiclana, F., Carter, J. and Fujita, H. (2016). A Hybrid Approach to the Sentiment Analysis Problem at the Sentence Level. *Knowledge-Based Systems*, 108: 110–124. <https://doi.org/10.1016/j.knosys.2016.05.040>.
- Ahuja R, Chug A, Kohli S, Gupta S, Ahuja P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Comput Sci.*, 152: 341–348.
- Akilandeswari J, Jothi G. (2018). Sentiment classification of tweets with non-language features. *Procedia Comput Sci.*, 143: 426–433
- Alessia D. (2015). Approaches, tools, and applications for sentiment analysis implementation. *Int Journal of Comput Appl.*, 125(3).

- Aydogan E. and M. A. Akcayol, (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques, *Int. Symp. Innov. Intell. Syst. Appl.*, pp. 17.
- Baecchi C., T. Uricchio, M. Bertini, and A. Del Bimbo, (2016). A multimodal feature learning approach for sentiment analysis of social network multimedia, *Multimed. Tools Appl.*, 75(5): 25072525.
- Bakker I, Van Der Voordt T, Vink P, De Boon J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Curr Psy-chol.*, 33(3): 405–421.
- Basiri M. E., Nemati S., Abdar M., Asadi S., & Acharry U. R. (2021). A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets. *Knowledge-Based Systems*. Journal homepage: www.elsevier.com/locate/knossys. Accepted 15 June 2022; Available online 25 June 2021.
- Bhaskar J, Sruthi K, Nedungadi P. (2015). Hybrid approach for emotion classification of audio conversation based on text and speech mining. *Procedia Comput Sci.*, 46: 635–643.
- Biltawi L M. (2016). Sentiment classification techniques for Arabic language a survey, *IEEE*.
- Binali, H. H.; Wu, C.; Potdar, V. (2009). A new significant area: Emotion detection in e-learning using opinion mining techniques. 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, pp. 259–264.
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges, and trends. *Knowledge-Based Systems*, 226: 107134.
- Cachola, I., Holgate, E., Preoțiu-Pietro, D., & Li, J. J. (2018). Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 2927-2938).
- Deng H., Daji E., Fangyao L., Ying C., & Bo Ma, (2022). Text sentiment analysis of fusion model based on attention mechanism. *The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021)*. *Procedia Computer Sci.*, 199: 741–748. Available online at www.sciencedirect.com.
- Dey, N., Borah, S., and Ashour, A.S. (2019). *Social Network Analytics. Computational Research Methods and Techniques*. <https://ja.1lib.world/book/5541614/eb0b79>.
- Dixon T (2012) “Emotion”: the history of a keyword in crisis. *Emot Rev.*, 4(4): 338–344.
- El-Din DM. (2015). Online paper review analysis. *Int J Adv Comput Sci Appl.*, 6(9).
- Gaurav S. Chavan, Sagar Manjare, Parikshit Hegde, Amruta Sankhe, (2014). A Survey of Various Machine Learning Techniques for Text Classification. *International Journal of Engineering Trends and Technology*, 15(6).
- Ghanbari-Adivi F, Mosleh M. (2019) Text emotion detection in social networks using a novel ensemble classifier based on Parzen tree estimator (TPE). *Neural Comput Appl.*, 31(12): 8971–8983
- Goel A. (2016). Real time sentiment analysis of tweets using naive bayes, *IEEE*.
- Hasan, A., Moin, S., Karim, A., and Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical and Computational Applications*, 23: 11. <https://doi.org/10.3390/mca23010011>.
- Haenlein M. and A. M. Kaplan, (2010). An empirical analysis of attitudinal and behavioral reactions toward the abandonment of unprofitable customer relationships, *J. Relatsh. Mark.*, 9(4): 200228.

- Heredia B., T. M. Khoshgoftar, J. Prusa, and M. Crawford, (2016). Cross-Domain Sentiment Analysis: An Empirical Investigation, 2016 IEEE 17th Int. Conf. Inf. Reuse Integr., pp. 160165.
- Ibrahim Rouby, (2018). Performance evaluation of an adopted sentiment analysis model for Arabic comments from Facebook, pp. 992-1045.
- Jaspreet Singh, Gurvinder Singh, and Rajinder Singh, (2017). Optimization of sentiment analysis using machine learning classifiers. Hum. Cent. Comput. Inf. Sci., 7: 32. <https://doi.org/10.1186/s13673-017-0116-3>.
- Jiang, S.; Qi, J. (2016). Cognitive Detection of Multiple Discrete Emotions from Chinese Online Reviews. Data Science in Cyberspace (DSC), IEEE International Conference on IEEE, pp. 137-142.
- Kaushik L. (2013). Sentiment extraction from natural audio streams, IEEE <https://doi.org/10.1109/icassp.2013.6639321>.
- Kim S-M. (2004). Determining the sentiment of opinions, ACM Digital Library, <https://doi.org/10.3115/1220355.1220555>.
- Kratzwald B, Ilić S, Kraus M, Feuerriegel S, Prendinger H. (2018). Deep learning for affective computing: text-based emotion recognition in decision support. Decis Support Syst., 115: 24–35.
- Khan R., Rustam F., Kanwal K., Mehmood A., & Choi G.S. (2021). US Based COVID-19 Tweets Sentiment Analysis Using TextBlob and Supervised Machine Learning Algorithms. 2021 International Conference on Artificial Intelligence (ICAI), <https://doi.org/10.1109/ICAI52203.2021.9445207>.
- Li C., B. Xu, G. Wu, S. He, G. Tian, and H. Hao, (2014). Recursive deep learning for sentiment analysis over social data, Proc. - 2014 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. - Work. 2: 13881429.
- Luo F., C. Li, and Z. Cao, (2016). Affective-feature-based sentiment analysis using SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support. Coop. Work Des., pp. 276281.
- Maha Heikal, (2018). Sentiment Analysis of Arabic Tweets using Deep Learning, 4th International Conference on Arabic Computational Linguistics (ACLing 2018), November 17-19 2018, Dubai, United Arab Emirates.
- Medhat W. (2014). Sentiment analysis algorithms and applications a survey. Ain Shams Eng J (Elsevier B.V.), 5(4): 1093–1113.
- Mohammad S. (2009). Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: Conference on empirical methods in natural language processing, pp. 599–608.
- Mohammad, S.M., Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. Computational Intelligence, vol. 31, no. 2, pp. 301–326.
- Muhammad Z. A., Fazli S., Muhammad I., Fazal M K., Shahboddin S., Amir M., Peter C., Annamaria R.V. (2020). Performance Evaluation of Supervised Machine Learning Techniques for Efficient Detection of Emotions from Online Content. Department of Mathematics and Informatics, J. Selye University, Komarno 94501, Slovakia, csibap@ujs.sk & koczya@ujs.sk
- Naseem U., Razzak I., Khushi M., Eklund P. W., & Kim J. (2020). COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. IEEE Transactions On Computational Social Systems.

- Nagamanjula R. and A. Pethalakshmi, (2020). A novel framework based on bi objective optimization and LAN2FIS for Twitter sentiment analysis. *Social Network Analysis and Mining*, 10:34 <https://doi.org/10.1007/s13278-020-00648-5>.
- Nagarajan SM, Gandhi UD. (2019). Classifying streaming of twitter data based on sentiment analysis using hybridization. *Neural Comput Appl.*, 31(5): 1425–1433.
- Neogi A. S., Garg K. A., Mishra R. K., & Dwivedi Y. K. (2021). Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights.*, 1(2021): 100019. Journal Homepage: www.elsevier.com/locate/jjime.
- Nofer, M. and Hinz, O. (2015). Using Twitter to Predict the Stock Market. *Business & Information Systems Engineering*, 57: 229–242. <https://doi.org/10.1007/s12599-015-0390-4>.
- Pang Bo and Lillian Lee, Shivakumar Vaithyanathan. (2002). “Thumbs up? Sentiment Classification using Machine Learning Techniques”. *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*.
- Pansy N. and Rupali V. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11: 81. <https://doi.org/10.1007/s13278-021-00776-6>.
- Phan, H. T., Tran, V. C., Nguyen, N. T., & Hwang, D. (2020). Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model. *IEEE Access*, 8, 14630-14641.
- Ray B., Avishek G., & Ram S. (2020). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing Journal*.
- Saif, M.M. (April 2017). Challenges in Sentiment Analysis. <https://www.researchgate.net/publication/316049100>.
- Semih Y. (2014). Tagging accuracy analysis on part-of-speech taggers. *J Comput Commun.*, 2: 157–162. <https://doi.org/10.4236/jcc.2014.24021>
- Severyn A. and A. Moschitti, (2015). Twitter Sentiment Analysis with Deep Convolutional Neural Networks, *Proc. 38th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR 15*, pp. 959-962.
- Shaila, S. G., Vadivel, A. (2015). Cognitive based sentence level emotion estimation through emotional expressions. *Progress in Systems Engineering*, pp. 707–713.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1631–1642.
- Sun S, Luo C, Chen J. (2017). A review of natural language processing techniques for opinion mining systems. *Inf Fusion.*, 36: 10–25.
- Sun, R., Wilson, N., Lynch, M. (2016). Emotion: A unified mechanistic interpretation from a cognitive architecture. *Cognitive Computation*, 8(1): 1–14.
- Strimaitis R., Stefanovič P., Ramanauskaitė S., & Slotkiene A. (2021). Financial Context News Sentiment Analysis for the Lithuanian Language. *Appl. Sci.*, 11: 4443. <https://doi.org/10.3390/app1110444>.

Sudhir P., & Suresh V. D. (2021). Comparative Study of Various Approaches, Applications and Classifiers for Sentiment Analysis, Global Transitions Proceedings (2021). <https://doi.org/10.1016/j.gltp.2021.08.004> .

Sweidan A. H., El-Bendary N., & Al-Feel H. (2021). Sentence-Level Aspect-Based Sentiment Analysis for Classifying Adverse Drug Reactions (ADRs) Using Hybrid Ontology-XLNet Transfer Learning. IEEE Access. <https://doi.org/10.1109/ACCESS.2021.3091394>.

Symeonidis S, Effrosynidis D, Arampatzis A. (2018). A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. Expert Syst Appl., 110: 298–310.

Umar Farooq, Hasan Mansoor, Antoine Nongaillard, Yacine Ouzrout, Muhammad Abdul Qadir, (2017). Negation Handling in Sentiment Analysis at Sentence Level, The Journal of Computers, 12(5): 470478.

Vaghela VB. (2016). Analysis of various sentiment classification techniques. Int J Comput Appl., 140(3).

Vateekul P. and T. Koomsubha, (2016). A Study of Sentiment Analysis Using Deep Learning Techniques on Thai Twitter Data.

Villavicencio C., Macrohon, J.J., Inbaraj, X.A., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes. Information, 12: 204. <https://doi.org/10.3390/info12050204>.

Wang Y., Huang G., Li J., LI H., Zhou Y., & Jiang H. (2021). Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis. IEEE Access. <https://doi.org/10.1109/ACCESS.2021.3062654>.

Winarsih, N. A. S.; Supriyanto, C. (2016). Evaluation of classification methods for Indonesian text emotion detection. International Seminar on Application for Technology of Information and Communication, pp. 130–133.

Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu, (2013). Unsupervised Sentiment Analysis with Emotional Signals. Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media. WWW, 2013 Rio de Janeiro, Brazil. ACM 978-1-4503-2035-1/13/05.

Yadav, A., and Vishwakarma, D.K. (2020). Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6): 4335–4385.

Yanmei L. and C. Yuda, (2015). Research on Chinese Micro-Blog Sentiment Analysis Based on Deep Learning, 2015 8th Int. Symp. Comput. Intell. Des., pp. 358361.

Yeole A. V., P. V. Chavan, and M. C. Nikose, (2015). Opinion mining for emotions determination, ICIIECS 2015 - 2015 IEEE Int. Conf. Innov. Information, Embed. Commun. Syst.